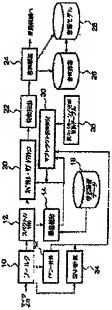
VOICE RECOGNITION DEVICE

Publication number:	JP2001215992
Publication date:	2001-08-10
Inventor:	AOSHIMA SHIGEKI
Applicant:	TOYOTA MOTOR CORP
Classification:	
- international:	***G10L15/20; G10L15/02; G10L21/02;*** G10L15/00; G10L21/00; (IPC1-7): G10L15/20; G10L15/02; G10L21/02; G10L101/02
- European:	
Application number:	JP20000022696 20000131
Priority number(s):	JP20000022696 20000131

Report a data error here

Abstract of JP2001215992

PROBLEM TO BE SOLVED: To provide a voice recognition device which surely recognizes inputted voice under various environmental conditions. SOLUTION: Inputted voice is supplied to a spectrum subtracting section 20 through a filter 10 and a spectrum analysis section 12. In the section 20, noise is subtracted from the inputted voice and the result is supplied to a feature extracting section 22. A noise difference section 14 computes the difference between input noise and the noise generated while learning a voice dictionary 26. The section 20 cancels the difference between the input noise and the noise of the dictionary 26 by subtracting the difference from the inputted voice. A subtracting magnification used in the subtraction is determined based on the SNR of the inputted voice and the difference spectrum from the section 14. The magnification is also determined for every analysis frame.



Data supplied from the *esp@cenet* database - Worldwide

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）

(11)特許出願公開番号
特開2001－215992
（P2001－215992A）

(43)公開日　平成13年8月10日(2001.8.10)

(51)Int.Cl.⁷	識別記号	ＦＩ		テーマコード゛(参考)
Ｇ１０Ｌ　15/20		Ｇ１０Ｌ 101:02		５Ｄ０１５
21/02		3/02	３０１Ｄ	９Ａ００１
15/02		7/08	Ａ	
// Ｇ１０Ｌ 101:02				

審査請求　未請求　請求項の数8　ＯＬ　（全 7 頁）

(21)出願番号　特願2000－22696(P2000－22696)

(22)出願日　平成12年1月31日(2000.1.31)

(71)出願人　000003207
トヨタ自動車株式会社
愛知県豊田市トヨタ町１番地

(72)発明者　青島　滋樹
愛知県豊田市トヨタ町１番地　トヨタ自動車株式会社内

(74)代理人　100075258
弁理士　吉田　研二　（外２名）

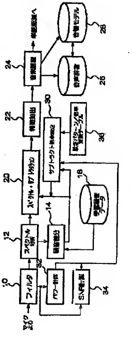
Ｆターム(参考)　5D015 EE05 FF04
9A001 HH05 HH17

(54)【発明の名称】　音声認識装置

(57)【要約】

【課題】　種々の環境下において確実に入力音声を認識する。

【解決手段】　入力音声はフィルタ１０及びスペクトル分析部１２を介してスペクトルサブトラクション部２０に供給される。スペクトルサブトラクション部２０では、入力音声から騒音を差し引き、特徴抽出部２２に供給する。騒音差分部１４では、入力騒音と音声辞書２６を学習させたときの騒音との相違を算出し、スペクトルサブトラクション部２０では入力音声からその相違分だけ差し引くことにより入力騒音と音声辞書２６の騒音との相違をキャンセルする。差し引く場合のサブトラクト倍率は、入力音声のＳＮＲや騒音差分部１４からの相違のスペクトルに基づき決定される。サブトラクト倍率は、分析フレーム毎に決定することもできる。



【特許請求の範囲】

【請求項１】　入力音声から騒音を差し引いて得られる音声の特徴を学習により得られた標準音声と比較して認識する音声認識装置であって、

前記標準音声に含まれる学習騒音と前記入力音声に含まれる入力騒音との相違に基づいて、前記入力音声から差し引くべき前記騒音を演算する演算手段と、

を有することを特徴とする音声認識装置。

【請求項２】　入力音声と学習により得られた標準音声とを比較することにより認識する音声認識装置であって、

前記標準音声に含まれる学習騒音と入力音声に含まれる入力騒音との相違に基づいて、前記標準音声に加算すべき騒音を演算する演算手段と、

を有することを特徴とする音声認識装置。

【請求項３】　請求項１、２のいずれかに記載の装置において、さらに、

前記入力音声のＳＮＲに応じて差し引くべき割合、あるいは加算すべき割合を決定する手段と、

を有することを特徴とする音声認識装置。

【請求項４】　請求項３記載の装置において、

前記入力音声のＳＮＲは、周波数領域での重み付けに基づいて算出されることを特徴とする音声認識装置。

【請求項５】　請求項１、２のいずれかに記載の装置において、さらに、

前記相違のスペクトル帯域毎に差し引くべき割合、あるいは加算すべき割合を決定する手段と、

を有することを特徴とする音声認識装置。

【請求項６】　請求項１、２のいずれかに記載の装置において、さらに、

前記入力騒音のパワー分散に応じて差し引くべき割合、あるいは加算すべき割合を決定する手段と、

を有することを特徴とする音声認識装置。

【請求項７】　請求項１、２のいずれかに記載の装置において、

前記入力騒音の音声分析フレーム毎のＳＮＲに基づいて差し引くべき割合、あるいは加算すべき割合を決定することを特徴とする音声認識装置。

【請求項８】　請求項１、２のいずれかに記載の装置において、

前記入力騒音の音声分析フレーム毎のパワーに基づいて差し引くべき割合、あるいは加算すべき割合を決定することを特徴とする音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は音声認識装置、特に騒音下において発生された音声を認識する技術に関する。

【０００２】

【従来の技術】従来より、入力音声から騒音を差し引いて得られる音声の特徴と予め学習により得られた標準音声とを比較することにより騒音下においても音声を認識する技術が知られている。

【０００３】たとえば、特開平１１−１５４０００号公報に開示された雑音抑圧装置及び該装置を用いた音声認識システムには、音声区間の入力信号に基づいて算出したパワースペクトルから雑音パワースペクトルに所定のサブトラクト係数を乗じたものを引き算することにより雑音の影響を排除して音声認識を行う技術が記載されている。

【０００４】

【発明が解決しようとする課題】一般に、雑音スペクトルを差し引くスペクトルサブトラクション技術においては、発生前の騒音区間の数十フレームを平均化することで騒音を推定し、この推定した騒音を音声区間の入力からフレーム毎（分析単位）に周波数領域で引き算するものである。

【０００５】しかしながら、このようにして騒音の影響を除去した入力音声と予め用意した標準パターンとを比較する場合、標準パターン（音声辞書）としてある程度の騒音が存在する環境下で発生した音声を用いる場合（無騒音に制御しても、完全には除去できないためある程度の騒音は残存する）には、比較の対象が騒音付の音声であるため、両者に相違が生じ、認識率が低下するおそれがある。

【０００６】また、上記従来技術においては、サブトラクト倍率を１より大きな値に設定しているが、これは推定騒音が平均化されているのに対して、パワーの大きい区間の音声に調整した場合の方が全体として認識率がよくなることを考慮したものであり、パワーが小さい区間においても同様にサブトラクト倍率を大きくすると騒音の引きすぎによる歪みが生じ、認識率が低下する問題もある。

【０００７】本発明は、上記従来技術の有する課題に鑑みなされたものであり、その目的は、比較すべき標準パターンが騒音下で発声されたパターンであっても確実に入力音声を認識することができ、また、種々の環境下においても認識率の低下を抑制することができる装置を提供することにある。

【０００８】

【課題を解決するための手段】上記目的を達成するために、本発明は、入力音声から騒音を差し引いて得られる音声の特徴を学習により得られた標準音声と比較して認識する音声認識装置であって、前記標準音声に含まれる学習騒音と前記入力音声に含まれる入力騒音との相違に基づいて、前記入力音声から差し引くべき前記騒音を演算する演算手段とを有することを特徴とする。学習時に含まれる騒音も考慮して差分演算することで、認識率の低下を有効に抑制できる。

【０００９】また、本発明は、入力音声と学習により得

られた標準音声とを比較することにより認識する音声認識装置であって、前記標準音声に含まれる学習騒音と入力音声に含まれる入力騒音との相違に基づいて、前記標準音声に加算すべき騒音を演算する演算手段とを有することを特徴とする。学習時に含まれる騒音も考慮して加算演算することで、認識率の低下を有効に抑制できる。

【００１０】ここで、前記入力音声のＳＮＲに応じて差し引くべき割合、あるいは加算すべき割合を決定する手段をさらに有することが好適である。雑音レベルが増大すると発声レベルも騒音レベルに比例して増大するランバード効果が存在するため、音声レベル（音声パワー）のみならず騒音レベル（騒音パワー）も考慮したＳＮＲで差し引くべき割合や加算割合を決定することで、特に音声パワーの大小によらず認識率を向上させることができる。ここで、ＳＮＲは音声パワーと騒音パワーの比で定義される。

【００１１】また、前記入力音声のＳＮＲは、周波数領域での重み付けに基づいて算出されることが好適であり、より具体的には人間の聴覚特性に基づいたフィルタ処理を行うことが望ましい。

【００１２】また、前記相違のスペクトル帯域毎、あるいは入力騒音のパワー分散に応じて差し引くべき割合、あるいは加算すべき割合を決定する手段をさらに有することが好適である。スペクトル帯域毎に割合を変化させることで、全ての帯域において認識率を向上させることができ、入力騒音のパワー分散に応じて割合を決定することで、ランバード効果を利用して認識率を向上させることができる。

【００１３】また、前記入力騒音の音声分析フレーム毎のＳＮＲあるいはパワーに基づいて割合を決定することも好適である。分析単位（フレーム）毎に騒音は変化するから、分析単位で割合を変化させることで、より高精度の認識が可能となる。

【００１４】

【発明の実施の形態】以下、図面に基づき本発明の実施形態について説明する。

【００１５】図１には、本実施形態の全体構成ブロック図が示されている。マイクから入力された入力音声はフィルタ１０を介してスペクトル分析部１２に供給される。なお、音声が入力されない場合には、騒音がフィルタ１０を介してスペクトル分析部１２に供給される。フィルタ１０は人間の聴覚特性を考慮したフィルタであり、具体的には周波数の高い領域を優先的に透過するフィルタである。フィルタ１０は必ずしも必須ではなく、マイクから入力された音声あるいは騒音を直接スペクトル分析部１２に供給してもよい。

【００１６】スペクトル分析部１２では、入力した音声や騒音をＦＦＴ等によりスペクトル分析し、周波数毎のパワーを算出する。算出されたスペクトルは平滑化され、騒音差分部１４に供給される。

【００１７】騒音差分部１４には、スペクトル分析部１２からの入力騒音スペクトル（マイクから音声が入力されず、騒音が入力された区間におけるスペクトルであり、音声に含まれる騒音と推定されるスペクトル）が供給されるとともに、比較の対象となる学習音声辞書の発声時に含まれていた学習騒音データを格納するデータベース１８から学習騒音スペクトルが供給される。騒音差分部１４では、これら２つのスペクトル、すなわち入力騒音スペクトルと学習騒音スペクトルとの差分を算出し、推定騒音スペクトルとする。具体的には、推定騒音スペクトルのＳＮＲ（推定騒音）は、

【数１】

$$\mathrm{SNR}（推定騒音）=\mathrm{SNR2}-\mathrm{SNR1} \quad \cdots (1)$$

で算出される。ただし、ＳＮＲ１は学習騒音スペクトルのＳＮＲであり、ＳＮＲ２は入力騒音スペクトルのＳＮＲである。ここで、ＳＮＲは、音声区間のパワーと騒音区間のパワーの比（Speech to Noise Ratio）として定義され、具体的には

【数２】

$$\mathrm{SNR}=10\mathrm{Log}(\Sigma P(Si)/\Sigma(i))/(\Sigma P(Nj)/\Sigma(i)) \quad \cdots (2)$$

で定義される。入力騒音スペクトルのＳＮＲは、スペクトル分析部１２で分析して得られた騒音のパワーと、発声実験値により得られた音声パワーとの比から算出することができる。学習騒音スペクトルのＳＮＲも同様である。

【００１８】以上のようにして入力騒音スペクトルと学習騒音スペクトルとの差分を演算することで両スペクトルの相違が演算されると、演算結果はスペクトルサブトラクション部２０に供給される。

【００１９】スペクトルサブトラクション部２０では、フィルタ１０及びスペクトル分析部１２を介して供給された入力音声パターン（音声区間における入力信号スペクトル）と騒音差分部１４から供給された推定騒音との差分を演算し、騒音の影響が除去された音声パターンを抽出して特徴抽出部２２に供給する。

【００２０】特徴抽出部２２は、騒音の影響が除去された入力音声パターンから特徴部分を抽出し、音素認識部２４に供給する。音素認識部２４では、予め学習により用意された音声辞書２６（この音声辞書の音声パターンには、学習時における騒音が付加されている）及び音響モデル２８に基づいて抽出された特徴がどの音素に該当するかを照合し、音素を認識して出力する。

【００２１】図２には、騒音差分部１４における差分演算が模式的に示されている。図において、（ａ）は学習

騒音スペクトルのＳＮＲ（ＳＮＲ１）が示されており、（ｂ）は入力騒音スペクトルのＳＮＲ（ＳＮＲ２）が示されている。騒音差分部１４では、供給されたこれら２つのＳＮＲに基づき、上述の（１）式に基づいてスペクトルサブトラクションすべき差分量を演算する。

【００２２】図3には、スペクトルサブトラクション部２０における差分の様子が模式的に示されている。フィルタ１０及びスペクトル分析部１２を介して供給された入力音声スペクトル（図中実線）と騒音差分部１４から供給された推定騒音（ＳＮＲ２－ＳＮＲ１であり、図中一点鎖線）との差分が演算され、これにより学習時の騒音と音声入力時の騒音との相違がキャンセルされ、精度よく音声辞書２６に記録された音声データと照合することができる。

【００２３】なお、上述した処理は、入力音声から入力騒音を差し引き、差し引いて得られたものにさらに学習騒音を付加して音声辞書２６に記録された騒音付音声データと照合すると考えることもできる。すなわち、上述した処理を数式で表現すると、（入力音声）－｛（入力騒音）－（学習騒音）｝＝（入力音声）－（入力騒音）＋（学習騒音）であり、音声辞書に学習時の騒音が付加されていても、これにより学習時の騒音に影響されずに認識できることが理解されよう。

【００２４】一方、スペクトルサブトラクション部２０にて入力音声から推定騒音を差し引く場合、差し引く倍率であるサブトラクト倍率を固定とした場合には、上述したように種々の環境下において安定して認識率を向上させることが困難となる。具体的には、パワーが小さい区間でサブトラクト倍率が大きくなりすぎ、騒音の引きすぎによる歪みが生じて認識率低下を招くことになる。

【００２５】そこで、本実施形態においてはさらにサブトラクト倍率設定部３０を設け、騒音差分部１４から出力された推定騒音にサブトラクト倍率αを乗じてスペクトルサブトラクション部２０に供給している。

【００２６】サブトラクト倍率設定部３０は、基本的には入力音声のパワーに応じてサブトラクト倍率を動的に変更するものであるが、一般に、図4に示されるように騒音レベルが増大すると発声レベルも騒音レベルにほぼ比例して増大する、いわゆるランバード効果が存在するため、最適のサブトラクト倍率を設定することは困難となる。そこで、本実施形態においては、図1に示されるようにフィルタ１０で高周波強調された入力音声のＳＮＲをＳＮＲ計算部３４で算出し、算出したＳＮＲをサブトラクト倍率設定部３０に供給し、サブトラクト倍率設定部３０で入力音声のＳＮＲに基づきサブトラクト倍率を設定している。具体的には、入力音声のＳＮＲが大きいほどサブトラクト倍率を大きく設定する。単に入力音声のパワーに応じてサブトラクト倍率を変更するのではなく、入力音声のＳＮＲに応じてサブトラクト倍率を変更することで、ランバード効果も考慮した高精度の音声認識が可能となり、特に入力音声のパワーが小さい区間における引きすぎを確実に防止できる。

【００２７】また、騒音が含まれていても、認識率が大きく低下する帯域と劣化の度合いが比較的小さい帯域が存在することが知られている。すなわち、騒音に強い帯域と弱い帯域が存在する。例えば、本願出願人は、１ｋＨｚ～３ｋＨｚに騒音スペクトルが存在すると、他の帯域に存在する場合に比べて認識率の低下が大きいことを確認している。したがって、ハイパスフィルタやローパスフィルタ等を用いて入力音声パターンから特定の帯域のみの信号を取り出して音声認識することにより、騒音環境下においても高精度に音声認識することが可能となる。しかしながら、騒音のスペクトルやパワーは種々変化するため、固定的な帯域通過フィルタ等を用いて音声認識する構成では、環境変化に柔軟に対応することができず、全体として見た場合に認識率の低下を招くおそれがある。

【００２８】そこで、本実施形態においては帯域毎にサブトラクション倍率を変化させ、種々の走行環境に柔軟に対応している。このため、図1に示されるように、騒音差分部１４から出力された推定騒音がサブトラクト倍率設定部３０に供給され、サブトラクト倍率設定部３０では、騒音パターン／倍率変換テーブル３６に基づいて推定騒音のスペクトル帯域毎にサブトラクト倍率を決定してスペクトルサブトラクション部２０で差し引くべき差分量を決定している。騒音パターン／倍率変換テーブル３６は、騒音パターンとその時の帯域毎のサブトラクト倍率を予め決定してテーブル形式で保持するもので、例えば、１ｋＨｚ～３ｋＨｚにおけるサブトラクト倍率を他の帯域に比べて大きくするように設定する。

【００２９】図5には、サブトラクト倍率設定部３０における処理が模式的に示されている。（ａ）及び（ｃ）は騒音差分部１４から出力された推定騒音のスペクトル例であり、（ａ）は比較的平坦なスペクトル、（ｃ）は低周波側に多くパワーが存在するスペクトル例である。（ｂ）は（ａ）が入力された場合に帯域毎に決定されるサブトラクト倍率であり、（ｄ）は（ｃ）が入力された場合の各帯域毎に決定されるサブトラクト倍率である。基本的には推定騒音のパワーに応じてサブトラクト倍率を変えているが（すなわちパワーが大なるほどサブトラクト倍率を増大させる）、さらに騒音に対し比較的認識率が低下しやすい帯域に対してはサブトラクト倍率を増大させている。このように、推定騒音、すなわち入力騒音スペクトルと学習騒音との差異のスペクトル帯域毎にサブトラクト倍率を決定することで、任意の走行環境、すなわち任意の騒音パターンに対しても高精度に認識することができる。

【００３０】なお、帯域毎のサブトラクト倍率αｉは、具体的には

【数3】

$$\alpha i=\beta i\cdot Pi \qquad \cdots (3)$$

で決定することができる。ここで、βｉは実験的に求めた帯域ｉの係数であり、Ｐｉは帯域ｉの推定騒音パワー、ｉは周波数帯域である。

【００３１】さらに、本実施形態においては図１に示されるようにフィルタ１０で高域強調された入力騒音の平均パワー及びその分散（あるいは偏差）をパワー計算部３２で算出し、サブトラクト倍率設定部３０に供給する構成となっている。サブトラクト倍率設定部３０では、パワーピーク値の平均値からの偏差、すなわちパワー分散値によりサブトラクト倍率を決定する。分散が大なるほどサブトラクト倍率を大きく設定し、分散が小なるほどサブトラクト倍率を小さく設定する。

【００３２】図６には、入力騒音のパワースペクトルと偏差の関係が示されている。図において、点線は入力騒音パワーの時間平均値であり、σ１及びσ２はピーク値の平均値からの偏差を示している。σ１＞σ２であり、偏差σ１の場合のサブトラクト倍率を偏差σ２の場合のサブトラクト倍率よりも大きく設定する。これにより、入力騒音パワーが少ない場合に発声レベルも少ないランバード効果が生じてもサブトラクト倍率が不必要に大きくなって騒音の引きすぎによる歪みが生じることがなく、認識率を向上させることができる。

【００３３】なお、上記実施形態においては、発声区間全体にわたってサブトラクト倍率を決定する場合について示したが、音声認識の分析フレーム単位でサブトラクト倍率を決定することも好適である。たとえば、マイクを２入力とし、１つの入力からの信号を用いて分析フレーム毎のＳＮＲを算出する。そして、このフレーム単位のＳＮＲに基づき、サブトラクト倍率を決定する。これにより、分析単位でのサブトラクト倍率設定が可能となり、音声認識率をより向上させることができる。もちろん、分析フレーム毎にサブトラクト倍率を決定する場合、入力騒音と学習騒音の相違を分析フレーム単位で算出し、このＳＮＲに基づいて決定することも好適である。また、ＳＮＲの代わりに、分析フレーム毎のパワーに基づいて倍率を変化させることも好適である。

【００３４】以上、本発明の実施形態について、入力音声から騒音を差し引いて得られる音声の特徴を音声辞書と比較する場合について説明したが、入力騒音と学習騒音との相違を算出し、音声辞書２６内のデータに加算して入力音声と比較することも可能であり、両者は技術的に等価である。そして、音声辞書２６に相違のデータを加算する場合の倍率もサブトラクト倍率と同様にＳＮＲやパワーに基づいて決定することができる。

【００３５】この場合の構成ブロック図が図７に示されている。図１と異なる点は、騒音差分部１４で算出した推定騒音をスペクトルアディション部２１に供給し、スペクトルアディション部２１では音声辞書２６に記憶された学習音声データにこの推定騒音、すなわち入力騒音と学習騒音の相違を付加する点である。なお、音声辞書２６の音声データに付加する際の倍率、すなわちアディション倍率はアディション倍率設定部３１で決定され（図１のサブトラクト倍率設定部３０に相当する）、アディション倍率設定部３１は、具体的には入力音声のＳＮＲやパワー分散、あるいは推定騒音のスペクトル帯域毎に倍率を決定する。

【００３６】

【発明の効果】以上説明したように、本発明によれば騒音環境下で標準音声を学習した場合においても、確実に入力音声を認識することができる。また、騒音が種々変化する任意の走行環境下において、走行認識率の低下を抑制することができる。

【図面の簡単な説明】

【図１】　実施形態の構成ブロック図である。

【図２】　騒音差分の処理説明図である。

【図３】　スペクトルサブトラクション説明図である。

【図４】　ランバード効果を示す説明図である。

【図５】　スペクトル帯域毎のサブトラクト倍率決定説明図である。

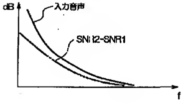
【図６】　入力音声パワーの分散を示すグラフ図である。

【図７】　他の実施形態の構成ブロック図である。

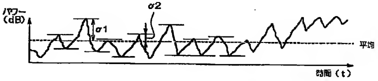
【符号の説明】

１０　フィルタ、１２　スペクトル分析部、１４　騒音差分部、１８　学習騒音データベース、２０　スペクトルサブトラクション部、２２　特徴抽出部、２４　音素認識部、２６　音声辞書、２８　音響モデルデータベース、３０　サブトラクト倍率設定部、３２　パワー計算部、３４　ＳＮＲ計算部、３６　騒音パターン／倍率変換テーブル。

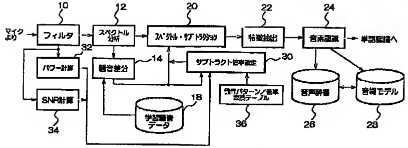
【図3】



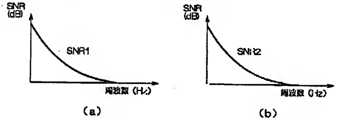
【図6】



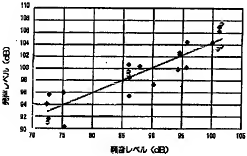
【図1】


マイクより
フィルタ
スペクトル分析
スペクトル・サブトラクション
特徴抽出
音素認識
単語認識へ
パワー計算
騒音差分
サブトラクト倍率設定
SNR計算
学習騒音データ
騒音パターン／倍率変換テーブル
音声辞書
音韻モデル
10
12
20
22
24
32
14
30
18
34
36
26
28

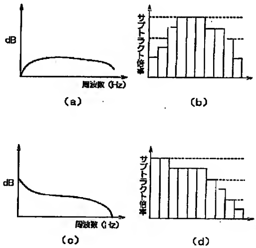
【図2】


SNR (dB)
SNR1
周波数 (Hz)
SNR2
(a)
(b)

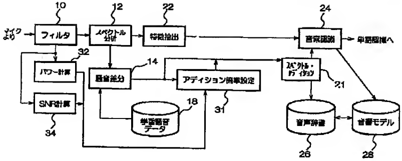
【図4】


110
108
106
104
102
100
98
96
94
92
90
発声レベル (dB)
70
75
80
85
90
95
100
105
騒音レベル (dB)

【図5】


dB
周波数 (Hz)
(a)
サブトラクト倍率
(b)
(c)
(d)

【図7】


マイクより
フィルタ
10
スペクトル分析
12
特徴抽出
22
音声認識
24
単語照合へ
パワー計算
32
SNR計算
34
雑音差分
14
学習雑音データ
18
アディション倍率設定
31
スペクトル・アディション
21
音声辞書
26
音響モデル
28